Project Management Web-Application with Data Analytics

MS Project Report

By

Sourabh Deshkulkarni

Rochester Institute of Technology

Department-Computer Science

Advisor: Dr. Leonid Reznik

December 1, 2016
# Table of Contents

- MS PROJECT REPORT ........................................................................................................................................... 1
- ADVISOR: DR. LEONID REZNIK ...................................................................................................................................... 1
- ABSTRACT ........................................................................................................................................................................... 3
- BACKGROUND ...................................................................................................................................................................... 4
- OBJECTIVES .......................................................................................................................................................................... 4
- TOOLS AND TECHNOLOGIES ............................................................................................................................................... 5
- FRAMEWORKS – SPRING 3.0 MVC, ANGULAR JS, D3 CHARTS .................................................................................................. 5
  - Application Architecture ....................................................................................................................................................... 5
  - MVC Architecture .................................................................................................................................................................... 6
- CORE IMPLEMENTATION ............................................................................................................................................................ 7
- DATABASE DESIGN AND IMPLEMENTATION .................................................................................................................................. 8
- REGISTRATION MODULE ............................................................................................................................................................ 10
- REGISTRATION FLOW DIAGRAM ............................................................................................................................................... 10
- REGISTRATION INTERNAL PROCESS ........................................................................................................................................ 10
- LOGIN MODULE ........................................................................................................................................................................ 11
- LOGIN FLOW DIAGRAM ............................................................................................................................................................ 11
- LOGIN INTERNAL PROCESS ......................................................................................................................................................... 11
- DATA ANALYTICS ....................................................................................................................................................................... 12
- DATA SET DETAILS ..................................................................................................................................................................... 13
  - Classification Model Results ................................................................................................................................................... 14
- PROJECT MANAGEMENT MODULE .......................................................................................................................................... 15
  - Project Upload ........................................................................................................................................................................... 16
  - Project Download ...................................................................................................................................................................... 16
  - Project Exploration .................................................................................................................................................................... 17
- RESULTS ..................................................................................................................................................................................... 18
  - Client Activity Analysis ........................................................................................................................................................... 18
  - Project Popularity Prediction .................................................................................................................................................... 20
- CHALLENGES FACED ................................................................................................................................................................. 22
- CONCLUSION ............................................................................................................................................................................. 23
- FUTURE WORK .......................................................................................................................................................................... 24
- REFERENCES .............................................................................................................................................................................. 25
- ATTACHMENT: APPLICATION MANUAL .................................................................................................................................. 26
Abstract

In today’s advancing world of Information Technology, it is imperative for a technocrat to manage the portfolio with the aid of all available platforms. Web platform facilitates cost-effective development, easy access from any device, anywhere and better storage along with security measures. This project leverages the benefits of web technologies and web platform to offer a project Management web application.

This project management application provides a platform to store and manage projects of any kind. The application tracks the client activity on the application and uses that to provide insights to the owner of the project. The insights include information such as most downloaded/ accessed project or any project file. The insights are not only limited to the present but the future predictions about project’s popularity are also offered with help of Data Analytics. The goal of this project is to leverage the blend of Web technologies and Data analytics to offer a smart solution to a technocrat to get analytical feedback on his work along with insights to improve the user experience of the application. The application holds a great potential and capacity to be used in many other similar contexts.
Background

Web technologies are improving everyday with better interactivity, high performance and cross compatibility with various tools and technologies. It is also observed that data mining as an emerging field is helping businesses and application owners to improve their product based on the historical data obtained from user activity and usage information. We can see blend of web applications and data mining in Market Surveillance and Customer Segmentation fields and the results of such analytics encourages the usage of this blend in other contexts.

[8] Paper talks about finding usage patterns from the web-based applications and making use of those patterns and insights to improve the experience for client. It also talked about different phases in determining those patterns, this project makes use of those steps/phases and builds on top of it to get client usage analysis and data analytics to predict the popular project along with project management utility.

Objectives

• Provide project management functionality through web application, as part of that offer project upload, download, explore and search functionality using Levenshtein’s string matching algorithm.
• Offer client activity analysis such as most downloaded/accessed project, most read/downloaded files.
• Provide project popularity prediction using Data Analytics using J48 classification algorithm with the aid of model built using client usage analytics.
• Provide optimum performance to the user and owner of the application in terms of project management and analytics using appropriate algorithms and data structures.
• With a blend of data mining and web technologies improve user experience and provide insights to the project owner.
**Tools and Technologies**

**Programming Languages** – Java, JavaScript, HTML, CSS, SQL, R  
**Frameworks** – Spring 3.0 MVC, Angular JS, D3 Charts  
**Tools** – Eclipse IDE, MySQL Workbench, Weka

---

**Application Architecture**

- As the pyramid indicates top of the web-application layer is User Interface which serves as the basic interaction.
- Angular JS glues the user interface with the application backend and serves as the UI decorator along with data population on UI [1] [3]
- Spring MVC serves as the backend which is responsible for the running the business logic and catering user requests. [2]
- MySQL stores the application specific data for Data analytics and Client activity analysis.
MVC Architecture

- MVC architecture helps in achieving loose coupling and high cohesion in the application.
- Different components of the application are divided into 3 distinct categories Model, View and Controller [7].
- All client requests are catered by the controller component, controller decides the flow of the request and builds the necessary business logic for each request [7].
- Model component holds the data and performs the encapsulation of the data. Model object is transported across the controller and view to pass the data between different components [7].
- View component is responsible to represent the response of the request. Controller sends the response data to the view [7].
- MVC architecture helps in bringing loose coupling into the project which helps in code reuse and application maintenance. [7]
Core Implementation

- Web application with Service Oriented Architecture is created with the aid of Spring MVC, Angular JS, MySQL and Java.
- Tomcat Apache container serves as the server for the application, Spring MVC REST endpoints are created to serve as a backend business logic.
- Angular JS supported user interface provides a functionality to perform all the application specific operations. [3]
- MySQL database stores the client activity data which is further used for the client activity analysis and project popularity prediction.
- Hence the web application provides three core functionalities that is Project management, Client activity analysis and Project popularity prediction.
The application makes use of MySQL 6.3.6 database to store the project-specific information. Information related to login- registration module, client activity data and project information. Below tables are being used to store the information –

1. userLogin
2. projectInfo
3. filesInfo

**userLogin**

This table holds the information about client login and registration. The table gets updated for every new registration and read for client login and authentication.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Data Type</th>
<th>Field Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>userId (pk)</td>
<td>VARCHAR(45)</td>
<td>Username of client.</td>
</tr>
<tr>
<td>password</td>
<td>VARCHAR(45)</td>
<td>Password of client.</td>
</tr>
<tr>
<td>name</td>
<td>VARCHAR(45)</td>
<td>Full name of client</td>
</tr>
<tr>
<td>registrationDate</td>
<td>DATETIME</td>
<td>Client registration date</td>
</tr>
<tr>
<td>lastActivity</td>
<td>DATETIME</td>
<td>Client last activity date</td>
</tr>
<tr>
<td>active_flag</td>
<td>INT(11)</td>
<td>Account active/inactive</td>
</tr>
<tr>
<td>city</td>
<td>VARCHAR(45)</td>
<td>Client address city</td>
</tr>
<tr>
<td>country</td>
<td>VARCHAR(45)</td>
<td>Client address country</td>
</tr>
<tr>
<td>gender</td>
<td>VARCHAR(45)</td>
<td>Client gender</td>
</tr>
</tbody>
</table>

The table is kept de-normalized as the current structure fulfills the requirement with necessary scaling scope. Also, the table would be used extensively for reading rather than writing. So, taking above facts into consideration current table structure for **userLogin** fulfills performance and functionality needs.
projectsInfo

This table holds the information about stored projects such as project name, number of files. This table is also responsible for storing client activity related to the project, such as project download count, project access count.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Data Type</th>
<th>Field Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>project_id (pk)</td>
<td>INT(11)</td>
<td>Unique id to reference project</td>
</tr>
<tr>
<td>Project_name</td>
<td>VARCHAR(45)</td>
<td>Name of the project</td>
</tr>
<tr>
<td>Project_Access_count</td>
<td>INT(11)</td>
<td>Project access count by client</td>
</tr>
<tr>
<td>Project_download_Count</td>
<td>INT(11)</td>
<td>Project download count by client</td>
</tr>
<tr>
<td>Access_Date_time</td>
<td>DATETIME</td>
<td>Project last access date</td>
</tr>
<tr>
<td>Project_technology</td>
<td>VARCHAR(45)</td>
<td>Prime technology used in the project</td>
</tr>
<tr>
<td>Project_type</td>
<td>VARCHAR(45)</td>
<td>Type of the project</td>
</tr>
<tr>
<td>numOfFiles</td>
<td>INT(11)</td>
<td>Total number of files in the project.</td>
</tr>
</tbody>
</table>

The table gets updated whenever a new project gets uploaded to the application as well whenever any project is accessed or downloaded. The table stores more miscellaneous data about project such type of the project, prime technology and number of files in the project. All this information serves as a core information for the data analytics part of the project to predict the most popular project.

filesInfo

This table holds the information about individual files from the project. Project_id acts as a foreign key to this table. The table also records file access and download count, which is in turn used for client activity analysis.

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Data Type</th>
<th>Field Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>file_id (pk)</td>
<td>INT(11)</td>
<td>Unique id to reference project</td>
</tr>
<tr>
<td>Project_id (fk)</td>
<td>INT(11)</td>
<td>Name of the project</td>
</tr>
<tr>
<td>File_name</td>
<td>VARCHAR(45)</td>
<td>Project access count by client</td>
</tr>
<tr>
<td>file_read_Count</td>
<td>INT(11)</td>
<td>Project download count by client</td>
</tr>
<tr>
<td>file_download_Count</td>
<td>INT(11)</td>
<td>Project last access date</td>
</tr>
<tr>
<td>Access_Date_time</td>
<td>DATETIME</td>
<td>Prime technology used in the project</td>
</tr>
</tbody>
</table>

This table gets updated when any new project is uploaded into the application and when individuals project files are accessed or downloaded from the application. The table has more write-intensive operations as it gets updated frequently to record the client activity.
Registration Module

This module is responsible for registering a new user into the application.

Registration flow diagram

1. User feeds the account credentials and personal information from the front end.
2. Angular JS controller bundles that information and creates a POST request to the backend server.
3. SPRING controller at the backend caters the POST request and establishes the connection with MySQL database.
4. User data is compared with existing data in database and if the entry exists then a warning is returned.
5. If entry doesn’t exist, then the entry is created for the new user.
6. The user password is encrypted at the client side and then bundled and sent as a payload to through the POST request.
7. The account information is later on used in login module.
Login Module

This module is responsible for validating a user credentials into the application.

Login flow diagram

Login Internal Process

1. User enters user-id and password on the index page.
2. Angular JS controller bundles that information and sends a POST request to the backend server.
3. SPRING controller at the backend caters the POST request and establishes the connection with MySQL database.
4. User data is compared with existing data in database and if the entry exists then the user is validated.
5. If entry doesn’t exist, then an error message is transported to the frontend.
6. If user is validated, then the application is redirected to the homepage.
7. The login information is sent to the server side through a POST request and with an encrypted password. The above measures are taken in order to improve the security of the application.
Data Analytics

This module is the core and most important part of the application. It predicts the popularity of the project based on various parameters. The dataset created for predicting the popularity is trained and prepared based on below assumptions –

1. Type of a project plays an important role and adds weightage to the project popularity.
2. Prime technology in the project is a pivotal factor, the project with popular and trending technology adds more weightage to project popularity.
3. Project access count, the project which is accessed most number of time is believed to be most downloaded and hence adds weightage to popularity.
4. Project which is most downloaded is believed to be most popular with the clients.
5. Project which is most recently accessed, that is within last 1 month could become popular.
6. Project with more number of files is assumed to have more to offer to the client/audiences, which might not be necessarily true all the time but for the scope of the project it is assumed so.

All of the above assumptions are devised based on real world experiences and some logical theories derived based on a discussion with the mentor. It is understood that the underlying parameters can be trained differently with inclusion of more project related attributes which in turn can change the model. Hence this model can potentially be tailored as per the need of the application.
Data Set Details

Classification techniques such as Random Forest, Naïve Bayes were also also tried out on the given dataset but below observations were recorded.

1. Naïve Bayes failed to select some of the important attributes while building the tree such as access count and project technology were missing from the tree.
2. Random Forest model was observed to be over-fitted with the training data even after 10-fold cross validation

It was observed that J48 provides better performance in terms of accuracy and attribute selection whilst building classification tree.

Dataset Size – 960 instances.
Number of attributes – 6
Classification algorithm – J48

1. numOfFiles – This attribute denotes number of unique files present in the project.
2. recentlyUsed – Boolean attribute denoting whether the project is accessed recently.
3. projectType – This attribute denotes the type of the project. Such as Machine learning, Distributed Systems etc.
4. primeTechnology – This attribute denotes the prime technology used in the project, such as Java, .NET etc.
5. accessCount – This attribute denotes the number of times the project is accessed.
6. downloadAccount – This attribute denotes the number of times the project is downloaded.

All above attributes are used to train the model to predict the popularity of the all uploaded projects in the application. They are weighted and trained using J48 classification algorithm. The foundational idea of J48 is to categorize the data into different ranges depending on the attribute values of the instances in the training data. J48 does classification based on the rules generated by above process or the decision tree [5].

Flow Diagram
1. Dataset for the Project popularity is created based on assumptions and experiments.
2. Dataset with around 1000 instances is trained using J48 classifier and 10-fold cross validation using Weka’s explorer.
3. All the information about the project such as project type, prime technology used in the project is recorded by the project owner while uploading the project and which is later on used for building the model.
4. Model is used in Java code to predict the popularity of the projects. [6]
5. WEKA’s Java API is used for predicting the value of unknown instance [6].
6. Prediction results are presented as part of a table and chart at the front end.
Project Management Module

Local Storage/Cloud

Unzip file

Spring file upload API

Angular JS Upload API

Download

Project Zip file

Spring file download API

Angular JS Download API

Download

Project Zip file
**Project Upload**

1. For project upload functionality, user is allowed to upload a zip project file up to 100 MB size.
2. Angular JS controller designated for Project upload functionality creates FormData object and appends multi-part file and other project attributes and attaches that to $http object for POST request.
3. Spring controller created backend URI receives the project file and unzips the file and stores on the local storage.
4. Multipart file is received at the backend and a utility to unzip the file is used to extract the files and directory from the project.

**Project Download**

1. A zip copy of a project is stored on the local storage along with the extracted copy.
2. Angular controller calls the GET request with project file name attached.
3. The GET request with a filename is catered by a Spring controller at the backend which searches for a project zip file and returns the zip file with content-type as ‘application’.
4. File with a zip extension and same name as uploaded gets downloaded on the disk.

The project upload and download functionality serves on cross-browsers and works as expected on Mac, Windows and Linux systems.
Project Exploration

1. Individual project can be explored through the web application.
2. Directory structure of the project along with all the files and preserved.
3. Individual file in the project can be read or downloaded as per the user choice.
4. Every directory inside the project can be explored recursively with the same structure as operating system explorer.
Results
Client Activity Analysis

Client activity is recorded in terms of accessed and downloaded files and projects. Hence the analysis is divided into 4 results [4]
1. Most read files from projects
2. Most downloaded files from projects
3. Most accessed projects
4. Most downloaded projects

It is believed that this analysis will provide insights to the project owner and help the project owner to understand the interests of the client and in turn maintain symbiotic relationship between the project owner and clients.
All the images attached below as part of the result are extracted from the application itself. [4]

Most Downloaded Files

![Pie chart showing the most downloaded files](image-url)
Most Read Files

- 3pr29yRAdTb.js
- 2Wa00kblH7.js
- 3evNwpKScgO.js
- 3O_t_oGmmUdf.css
- 2x5P32tRQ-.js

Most Downloaded Projects

- Dk6_new_test
- Dk5_new_test
- Dk2_new_test
- Dk_new_test
- Dk4_new_test
The results related to this module are divided into two parts –
1. Pie chart representing division of low, average and high popularity number of projects.
2. The second result is a table listing down the projects as per their popularity.
Hence the results provide client activity analysis and offers future prediction of project popularity using D3 charts framework which is compatible with the Angular JS framework and provides interactivity in terms of the user experience.

<table>
<thead>
<tr>
<th>Popularity</th>
<th>Project ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Dk2_new_test</td>
</tr>
<tr>
<td>High</td>
<td>Dk3_new_test</td>
</tr>
<tr>
<td>Average</td>
<td>Dk4_new_test</td>
</tr>
<tr>
<td>Average</td>
<td>Dk5_new_test</td>
</tr>
<tr>
<td>Average</td>
<td>Dk_new_test</td>
</tr>
</tbody>
</table>
Challenges faced

1. Angular JS doesn’t support data type as file, i.e. content-type – file, project upload/download being the core functionality of the application, choosing Angular JS as a front end framework was a tough choice but Angular JS provides a work-around for that with the help of directive which helps in overcoming the issue [1].

2. Being a single page application with SOA, it was difficult to save the previous state of the page.

3. Back button navigation had to be controlled for project exploration module of the application.

4. Integrating data analytics part was a challenge considering no availability of the historical data.

5. For data analytics part, customized data had to be prepared with some logical assumptions and future contexts.

6. Integrating Weka API into Java, Spring application made it bulky.

7. As an array of different technologies were used, it was a challenging task to find the right technologies which are compatible with each other and serve the objective of the application at the same time.
Conclusion

1. A web application with the blend of data analytics has been created without any compatibility issues.
2. Instant real time analysis of the data as well data mining has been achieved with the help of cutting edge technologies such as Java, Spring 3, Angular JS and Weka.
3. Project management with a maximum project size of 100 MB is successfully achieved with optimum performance and no delay.
4. J48 classification works well for the modeled dataset and generates accurate results.
5. Expected project management functionality along with data mining has been used to potentially improve client and application owner experience.
Future Work

1. Use of HTTPS and better security mechanisms to protect the application from security threats.
2. Use load balancing technologies to improve application reliability.
3. More data mining capabilities and contexts could be added.
4. Different user account roles and privileges could be added once application scope grows.
5. Cloud server could be used instead of local server to improve storage capacity.
References

[5] Decision Tree Analysis on J48 Algorithm for Data Mining by Manish Mathuria
[6] Integrating an Advanced Classifier in WEKA by Paul ,Stefan Popescu, Mihai Mocanu and Marian Cristian Mihaescu
Attachment:
Application Manual
Installation

1. Download and copy latest Tomcat Apache web server.
2. Download and install MySQL workbench and create the necessary tables.
3. Copy cloudsole-angular.war file in the web-apps folder of tomcat directory.
4. The war file is a deployable application web archive.
5. Open bin directory of Tomcat server and run startup.bat file to start the server.

Application End-point URL

1. Please open - http://localhost:8080/cloudsole-angular/ to access the application.
1. Please click on register and input all the details to create the account.
2. Once the account is created, you will be redirected to the login screen, please use the login credentials to login.

![Login Screen]

- dksourabh
- **********

Login »

Register »
Project Upload

Click on ‘Project Upload’ from the left hand menu list.

1. Select the .zip file to be uploaded.
2. Choose the project type from drop down.
3. Choose Project technology from drop down.
4. Click on upload to upload the project.

The project will be uploaded on the local drive with unzipped version.
Click on ‘Project Download’ from the left hand menu list.

1. All uploaded projects will be presented in this section for download.
2. Click on the project name in green to download the project.
3. Click on view to explore the projects.
4. Explore the files from the projects, click on read file to read the file, click on the file name to download the project.
5. If there are directories in the project, then click on view to view the directories.
6. The files and directories will be viewed and explored just as explored through the operating system explorer.
Click on ‘Data Analytics’ to view the Data analytics section.

Each project with popularity index will be presented along with the total project popularity pie chart. This is a future popularity prediction.
Client Usage Analysis

Click on ‘Client Activity Analysis’ to find the client usage analysis and insights about the same.
Application shutdown

Navigate to bin folder in Tomcat apache and click on shutwon.bat to stop the sever and application.